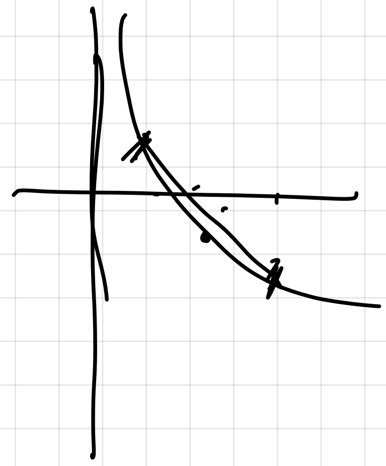
HW2. 2022-18599 이기탁

#2.

Points of non-diff: $Y_i^T X_i \theta = 1$.

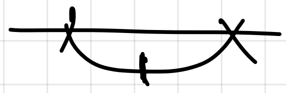
If this point is entered, I added it to a list and printed it. It turned out that this point was not entered.

#4. Lemma) $-\log x$ is convex. ($x \in (0, \infty)$).



pf) Take $x_1 < x_2$ and $\eta \in (0,1)$.

Let $f(x) = \frac{\log x_2 - \log x_1}{x_2 - x_1}(x - x_1) - (\log x - \log x_1)$



Then, $f(x_1) = 0$, $f(x_2) = 0$, and $f'(x) = \frac{\log x_2 - \log x_1}{x_2 - x_1} - \frac{1}{x}$.

$f''(x) = \frac{1}{x^2} \Rightarrow f'$ increases.

By MVT, $\exists c \in (x_1, x_2)$ s.t. $f'(c) = 0$.

For $y \in (x_1, c)$, $\exists z \in (x_1, y)$ s.t. $f'(z) = \frac{f(y)}{y - x_1} < 0$. Hence, $f(y) < 0$.

For $y \in (c, x_2)$, $\exists z \in (y, x_2)$ s.t. $f'(z) = \frac{-f(y)}{x_2 - y} > 0$. Hence, $f(y) < 0$.

Hence, $f(\eta x_1 + (1-\eta)x_2) < 0$ for all $\eta \in (0,1)$.

$$f(\eta x_1 + (1-\eta)x_2) = \underbrace{\frac{\log x_2 - \log x_1}{x_2 - x_1} \cdot (x - x_1) + \log x_1}_{= \, \eta \log x_1 + (1-\eta)\log x_2} - \log(\eta x_1 + (1-\eta) x_2)$$

$\therefore -\log x$ is convex. ☒

$$D_{KL}(p \| q) = \mathbb{E}_i\left[\log \frac{p_i}{q_i}\right] = \mathbb{E}_i\left[-\log \frac{q_i}{p_i}\right] \geq \log \mathbb{E}_i\left[\frac{q_i}{p_i}\right] = 0.$$ ☒

$\because \int \frac{q}{p} \cdot p \cdot d\mu = 1$.

~~#5~~. By Lemma of 4, $-\log x$ is strictly convex.

Say $X = \frac{p_i}{q_i}$. This is non-constant if $p \neq q$, because if

$\frac{p_i}{q_i} = c$, $\int q_i = \int p_i = c \int q_i$, so $c = 1$. Then $\underline{p = q}$.

$\therefore D_{KL}(p \| q) = \mathbb{E}_i[\log \frac{p}{q}] = \mathbb{E}[-\log \frac{q}{p}] > \log \mathbb{E}[\frac{q}{p}] = 0$.

#6.

$$f_\theta(x) = \sum_{j=1}^{p} u_j \cdot \sigma(a_j x + b_j).$$

$$\frac{\partial f}{\partial u_j} = \sigma(a_j x + b_j) \Rightarrow \nabla_u f = \sigma(ax + b).$$

$$\frac{\partial f}{\partial a_j} = u_j \cdot x \cdot \sigma'(a_j x + b_j) \Rightarrow \nabla_a f = \sigma'(ax+b) \odot ux$$

$$\frac{\partial f}{\partial b_j} = u_j \cdot \sigma'(a_j x + b_j) \Rightarrow \nabla_b f = \sigma'(ax+b) \odot u$$

$u \odot v = \mathrm{diag}(u) v$. because $\begin{pmatrix} u_1 & & \\ & \ddots & \\ & & u_p \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_p \end{pmatrix} = \begin{pmatrix} u_1 v_1 \\ \vdots \\ u_p v_p \end{pmatrix} = u \odot v$.

#7. It almost matches f^*.